The Rook on the Half-Chessboard, or How Not to Diagonalize a Matrix

Kiran S. Kedlaya
Lenhard L. Ng

1 Introduction

We show that the "reverse triangular" matrix

$$T_n = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 & 1 \\ 0 & 0 & \cdots & 0 & 1/2 & 1/2 \\ 0 & 0 & \cdots & 1/3 & 1/3 & 1/3 \\ \vdots & \vdots & ⋰ & \vdots & \vdots & \vdots \\ 0 & 1/(n-1) & \cdots & 1/(n-1) & 1/(n-1) & 1/(n-1) \\ 1/n & 1/n & \cdots & 1/n & 1/n & 1/n \end{pmatrix},$$

that is,

$$(T_n)_{ij} = \begin{cases} 1/i & \text{if } i+j \geq n+1 \\ 0 & \text{otherwise,} \end{cases}$$

has eigenvalues

$$1, -1/2, 1/3, \ldots, (-1)^{n+1}/n.$$

In Section 2, we establish this result in two ways: (1) by explicitly exhibiting eigenvectors via an odd-looking combinatorial identity, and (2) by giving a matrix that conjugates T_n into an upper triangular matrix.

In Section 3, we use the knowledge of the eigenvalues of T_n to study two related random walks, most conveniently summarized as "the rook on the half-chessboard." Consider an $n \times n$ chessboard from which all squares above (but not including) the northwest-southeast diagonal have been removed. In the "sloppy" random walk, a rook is moved as follows. If a coin flip comes up heads, we choose a square uniformly at random in the row containing the rook (including the square the rook occupies), and move the rook there. If the coin flip comes up tails, we do the same using the column instead of the row.

A fundamental property of this random walk is the speed at which an initial probability distribution for the position of the rook converges to the uniform distribution. The sloppy walk has an obvious inefficiency built in: two or more moves in the same direction have no more effect on the distribution than a single move in that direction. Hence we also introduce the "ordered" random walk, in which horizontal and vertical moves alternate. We might expect that the ordered walk converges twice as quickly as the sloppy walk, since on average only half of the moves in the sloppy walk have any effect; we discover that it actually converges somewhat more quickly than that.

2 How (Not) to Diagonalize a Matrix

We first establish a combinatorial identity (Proposition 1), of which we use only a special case (Corollary 2) for the diagonalization.

Proposition 1 *Let k, ℓ, m, n be nonnegative integers. Then*

$$\sum_j (-1)^j \tbinom{m+j}{m-1}\tbinom{n-j-1}{n-m}\tbinom{n-k-1}{j-k+\ell} = (-1)^{k+\ell+m+1}\sum_j (-1)^j \tbinom{m+j}{m}\tbinom{n-j-1}{n-m}\tbinom{\ell}{j-k+\ell}.$$

PROOF. For convenience, define

$$F(k,\ell,m,n) = (-1)^{k+\ell+m+1}\sum_j (-1)^j \binom{m+j}{m-1}\binom{n-j-1}{n-m}\binom{n-k-1}{j-k+\ell}$$

and

$$G(k,\ell,m,n) = \sum_j (-1)^j \binom{m+j}{m}\binom{n-j-1}{n-m}\binom{\ell}{j-k+\ell};$$

we claim that $F \equiv G$. Using the identities $\tbinom{n-k-1}{j-k+\ell} = \tbinom{n-k}{j-k+\ell} - \tbinom{n-k-1}{j-k+\ell-1}$ and $\tbinom{\ell}{j-k+\ell} = \tbinom{\ell-1}{j-k+\ell} + \tbinom{\ell-1}{j-k+\ell-1}$, we obtain

$$F(k,\ell,m,n) = F(k-1,\ell-1,m,n) + F(k,\ell-1,m,n),$$

$$G(k,\ell,m,n) = G(k-1,\ell-1,m,n) + G(k,\ell-1,m,n).$$

From these functional equations, we conclude that if $F(k,\ell-1,m,n) = G(k,\ell-1,m,n)$ for fixed ℓ and all k,m,n, then $F(k,\ell,m,n) = G(k,\ell,m,n)$ for all k,m,n as well. Hence the claim follows by induction, once we show $F(k,0,m,n) = G(k,0,m,n)$ for all k,m,n. Since the only nonzero term in $G(k,0,m,n)$ occurs at $j = k$, this becomes

$$\sum_j (-1)^j \tbinom{m+j}{m-1}\tbinom{n-j-1}{n-m}\tbinom{n-k-1}{j-k} = (-1)^{m+1}\tbinom{m+k}{m}\tbinom{n-k-1}{n-m}.$$

Upon expanding the binomial coefficients in terms of factorials, the terms involving n cancel out, leaving us to show that

$$\sum_j (-1)^j \binom{m+j}{j-k}\binom{m}{j+1} = (-1)^{m+1}\binom{m-1}{k}.$$

This can be verified either using the WZ method [3] with the certificate

$$\frac{(2m+1)(j+1)(k-j)}{m(j-m)(m+1+k)},$$

or directly using upper negation and the Vandermonde identity [2, p. 174]:

$$\begin{aligned}\sum_j (-1)^j \binom{m+j}{j-k}\binom{m}{j+1} &= (-1)^k \sum_j \binom{-k-m-1}{j-k}\binom{m}{m-j-1} \\ &= (-1)^k \binom{-k-1}{m-k-1} \\ &= (-1)^{m+1}\binom{m-1}{k}. \quad \square\end{aligned}$$

Corollary 2 *Let k, m, n be nonnegative integers. Then*

$$\sum_j (-1)^j \tbinom{m+j}{m-1}\tbinom{n-j-1}{n-m}\tbinom{n-k-1}{j} = (-1)^{m+1}\sum_j (-1)^j \tbinom{m+j}{m}\tbinom{n-j-1}{n-m}\tbinom{k}{j}.$$

Proof. Set $\ell = k$ in Proposition 1. □

Proposition 1 can be derived from Corollary 2 by induction on $\ell - k$, using the functional equations. This observation is what led us to formulate Proposition 1 in the first place.

Proposition 3 *The eigenvalues of T_n are $(-1)^{m+1}/m$ for $m = 1, 2, \ldots, n$.*

First Proof. Since the proposed eigenvalues for T_n are distinct, it suffices to exhibit an eigenvector for each. We claim the vector $(x_{n-1}, x_{n-2}, \ldots, x_1, x_0)$, with entries

$$x_i = \sum_{j=0}^{m-1} (-1)^j \binom{m+j}{m}\binom{n-j-1}{n-m}\binom{i}{j}, \qquad i = 0, 1, \ldots, n-1,$$

is an eigenvector of T_n (nonzero because $x_0 \neq 0$) with eigenvalue $(-1)^{m+1}/m$. It suffices to show that $(x_0 + x_1 + \cdots + x_{k-1})/k = (-1)^{m+1} x_{n-k}/m$; however, the latter follows from the identity $\sum_{i=0}^{k-1}\binom{i}{j} = \binom{k}{j+1}$ and Corollary 2, which may be rewritten (after substituting $n-k$ for k) as

$$m\sum_{j=0}^{m-1} (-1)^j \tbinom{m+j}{m}\tbinom{n-j-1}{n-m}\tbinom{k}{j+1} = (-1)^{m+1} k \sum_{j=0}^{m-1} (-1)^j \tbinom{m+j}{m}\tbinom{n-j-1}{n-m}\tbinom{n-k}{j}. \quad \square$$

Second Proof (outline). Let M be the $n \times n$ matrix given by

$$M_{ij} = \binom{n-i}{j-1}.$$

One can then check in succession that

$$(M^{-1})_{ij} = (-1)^{n+i+j+1}\binom{i-1}{n-j},$$

$$(T_nM)_{ij} = \frac{1}{i}\binom{i}{j},$$

and

$$(M^{-1}T_nM)_{ij} = \frac{(-1)^{i+1}}{j}\binom{n-i}{n-j}.$$

Each of these is a combinatorial identity that can be checked using either WZ or upper negation and Vandermonde. Since $M^{-1}T_nM$ is upper triangular, its diagonal entries are the eigenvalues of T_n. □

3 Application to Random Walks

The "sloppy" Markov process described in Section 1 is more precisely given by the (symmetric, doubly stochastic) transition matrix

$$K(v_{ij}, v_{i'j'}) = \begin{cases} 1/(2(n-i)) + 1/(2j) & \text{if } i = i' \text{ and } j = j' \\ 1/(2(n-i)) & \text{if } i = i' \text{ and } j \neq j' \\ 1/(2j) & \text{if } i \neq i' \text{ and } j = j' \\ 0 & \text{otherwise,} \end{cases}$$

where v_{ij} denotes the square in the $(i+1)$-st row and the j-th column (so that $0 \leq i < j \leq n$). Here $K(v_{ij}, v_{i'j'})$ denotes the probability of ending in $v_{i'j'}$ after one move, after beginning in v_{ij}.

To display the transition matrix K' for the "ordered" walk, first define the auxiliary matrices K_1 and K_2 as the transition matrices associated with horizontal and vertical moves, respectively; that is,

$$K_1(v_{ij}, v_{i'j'}) = \begin{cases} 1/j & \text{if } j = j' \\ 0 & \text{otherwise} \end{cases}$$

and

$$K_2(v_{ij}, v_{i'j'}) = \begin{cases} 1/(n-i) & \text{if } i = i' \\ 0 & \text{otherwise.} \end{cases}$$

Then $K' = K_2K_1$. (Note that $K = (K_1 + K_2)/2$.)

In order to calculate how quickly the sloppy and ordered random walks become random, we must first compute the eigenvalues of their transition matrices.

Proposition 4 *The eigenvalues of K are 0, 1, and $1/2 \pm 1/(2m)$, where $m = 2, \ldots, n$; each has multiplicity 1 except for 0, which has multiplicity $\binom{n-1}{2}$.*

Proof. We produce $\binom{n+1}{2} = \binom{n-1}{2} + 2(n-1) + 1$ independent eigenvectors with the asserted eigenvalues; by counting, these give all of the eigenvalues.

If $a, b \in \{1, \ldots, n-1\}$ and $a < b$, define f_{ab} by $f_{ab}(v_{ab}) = 1$, $f_{ab}(v_{a-1,b}) = -1$, $f_{ab}(v_{a,b+1}) = -1$, $f_{ab}(v_{a-1,b+1}) = 1$, and $f_{ab}(v_{ij}) = 0$ for all other v_{ij}. Then $K_1f_{ab} = K_2f_{ab} = 0$, so that $Kf_{ab} = 0$. There are $\binom{n-1}{2}$ ways to choose such pairs a, b, and it is straightforward to check that all f_{ab}'s thus produced are linearly independent; thus 0 is an eigenvalue of K with multiplicity (at least) $\binom{n-1}{2}$.

Next note that 1 is an eigenvalue of K with eigenvector $(1, 1, \ldots, 1)$.

Now let m be in $\{2, \ldots, n\}$. By Proposition 3, we know that T_n has an eigenvector $(x_{n-1}, x_{n-2}, \ldots, x_1, x_0)$ with eigenvalue $(-1)^{m+1}/m$. Define f_1 and f_2 by $f_1(v_{ij}) = x_{n-j}$ and $f_2(v_{ij}) = x_i$. Then

$$K_1 f_1(v_{ij}) = \sum_{i'<j'} K_1(v_{ij}, v_{i'j'}) f_1(v_{i'j'}) = \sum_{i'=0}^{j-1} \frac{f_1(v_{i'j})}{j} = f_1(v_{ij})$$

so that $K_1 f_1 = f_1$. Moreover,

$$K_2 f_1(v_{ij}) = \sum_{j'=i+1}^{n} \frac{f_1(v_{ij'})}{n-i} = \frac{1}{n-i} \sum_{j'=i+1}^{n} x_{n-j'} = \frac{(-1)^{m+1} x_i}{m}$$

by the eigenvector relation, so that $K_2 f_1 = (-1)^{m+1} f_2/m$. Similarly, $K_1 f_2 = (-1)^{m+1} f_1/m$ and $K_2 f_2 = f_2$. Hence if we define f_+ and f_- by $f_\pm = f_1 \pm f_2$, then

$$K f_\pm = \frac{1}{2}\left(1 \pm \frac{(-1)^{m+1}}{m}\right) f_\pm.$$

It is easy to check that $f_\pm \neq 0$; we conclude that $1/2 \pm 1/(2m)$ are eigenvalues of K for $m = 2, \ldots, n$. □

Proposition 5 *The eigenvalues of K' are 0 and $1/m^2, m = 1, 2, \ldots, n$; each has multiplicity 1 except for 0, which has multiplicity $\binom{n}{2}$.*

Proof. As in the proof of Proposition 4, we first find $\binom{n}{2}$ independent eigenvectors with eigenvalue 0. In this case, they are given by $f_{ab}, 1 \leq a < b \leq n$, defined by $f_{ab}(v_{ab}) = 1, f_{ab}(v_{a-1,b}) = -1$, and $f_{ab}(v_{ij}) = 0$ for all other v_{ij}. It is straightforward to check that these eigenvectors are linearly independent.

As before, 1 is an eigenvalue of K' with eigenvector (1,1,. . . ,1).

Now let m be in $\{2, \ldots, n\}$. Define f_1 and f_2 as in the proof of Proposition 4. Then

$$K' f_2 = K_2 K_1 f_2 = \frac{(-1)^{m+1}}{m} K_2 f_1 = \left(\frac{(-1)^{m+1}}{m}\right)^2 f_2 = \frac{1}{m^2} f_2,$$

and f_2 is an eigenvector of K' with eigenvalue $1/m^2$, as desired. □

Standard techniques applied to the spectra of K and K' give bounds on how quickly these Markov chains approach stationarity [1]. Roughly speaking, suppose that M is a diagonalizable stochastic transition matrix with eigenvalues $1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_q$ and corresponding eigenvectors $v_1 = (1, \ldots, 1), v_2, \ldots, v_q$. For any starting probability distribution $v = c_1 v_1 + c_2 v_2 + \cdots + c_q v_q$ (where $c_1 = 1/q$, since the sum of the entries in v must be 1), we have $M^r v - v_1/q = c_2\lambda_2^r v_2 + \cdots + c_q \lambda_q^r v_m$; thus the random walk described by M converges to stationarity at an approximate rate of λ_2^r, where r is the number of moves made.

More precisely, let μ be the uniform distribution on the state space $\{v_{ij}\}$, i.e., $\mu(v_{ij}) = 1/\binom{n+1}{2}$, and let $K^r_{v_{ij}}(v_{i'j'}) = K^r(v_{ij}, v_{i'j'})$ be the $(v_{ij}, v_{i'j'})$ entry of the matrix K^r, i.e., the probability that the chain is in state $v_{i'j'}$ after r steps, after beginning in state v_{ij}. A measure of how close K^r is to randomness is the total variation distance

$$\|K^r_{v_{ij}}\|_{\mathrm{TV}} = \frac{1}{2} \sum_{i'<j'} |(K^r_{v_{ij}} - \mu)(v_{i'j'})|,$$

and similarly for K'. Then the techniques of [1] give the following asymptotic bounds on the distance from randomness of the chains K and K' after r steps. (Here $\sim$ represents asymptotic behavior for large r.)

Proposition 6 *For any $i < j$, $\|K^r_{v_{ij}} - \mu\|_{\mathrm{TV}} \sim \frac{1}{2}\left(\frac{3}{4}\right)^r$.*

Proposition 7 *For any $i < j$, $\|K'^r_{v_{ij}} - \mu\|_{\mathrm{TV}} \sim \frac{1}{2}\left(\frac{1}{4}\right)^r$.*

In other words, it takes approximately $c/\log(4/3)$ steps for the sloppy walk to "become random," in the sense of being distance e^{-c} from uniformity, and $c/\log 4$ steps for the ordered walk to do so, more or less independently of n. Since one step of the ordered walk actually involves two moves, the ordered walk becomes random more quickly by a factor of

$$\frac{\log 4}{2\log(4/3)} \approx 2.409,$$

which is greater than the factor of 2 we initially suggested.

It would be interesting to extend these results to more than two dimensions; to avoid taxing the reader's visual faculties, we suggest an alternate formulation. The squares of the half-chessboard can be identified with the increasing integer sequences $0 \leq m_1 < m_2 \leq n$, where the sloppy walk proceeds by either moving m_1 between 0 and $m_2 - 1$ or moving m_2 between $m_1 + 1$ and n. The k-dimensional analogue of the sloppy walk would be to walk among the increasing integer sequences $0 \leq m_1 < \cdots < m_k \leq n$ by choosing $i \in \{1, \ldots, k\}$ uniformly at random, then moving m_i between $m_{i-1} + 1$ and $m_{i+1} - 1$ (where $m_0 = -1$ and $m_{k+1} = n + 1$). Our computations suggest that the eigenvalues of the resulting transition matrix are not as easily described as in the two-dimensional case; moreover, it is not even clear what the correct analogue of the ordered walk should be.

Acknowledgments

The authors would like to thank Persi Diaconis for proposing the random walk problem, one of the referees for noting the rook-on-a-half-chessboard interpretation of this random walk, and Manjul Bhargava and Eric Rains for suggesting the second proof of Proposition 3.

References

[1] P. Diaconis and L. Saloff-Coste, Comparison theorems for reversible Markov chains, *Ann. Appl. Prob.* **3** (1993), 696–730.

[2] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics*, Addison-Wesley, Reading, 1989.

[3] H.S. Wilf and D. Zeilberger, Rational functions certify combinatorial identities, *J. Amer. Math. Soc.* **3** (1990), 147–158.

Department of Mathematics
Massachusetts Institute of Technology
Cambridge, MA 02139
kedlaya@math.mit.edu

Department of Mathematics
Massachusetts Institute of Technology
Cambridge, MA 02139
lenny@math.mit.edu